Lung ultrasound vs chest radiography for pneumonia diagnosis in critically ill patients in the intensive care unit: a systematic review and Bayesian network diagnostic meta-analysis

Daniele Orso,[a,*] *Federico Fonda,*[a] *and Tiziana Bove*[a,b]

[a]Department of Emergency, Udine University Hospital, Azienda Sanitaria Universitaria Friuli Centrale (ASU FC), 33100, Udine, Italy
[b]Department of Medicine (DMED), University of Udine, 33100, Udine, Italy

eClinicalMedicine
2025;89: 103553

Published Online xxx
https://doi.org/10.1016/j.eclinm.2025.103553

Summary

Background Diagnosing pneumonia in intensive care unit (ICU) patients is challenging. Chest X-ray (CXR) is widely used, whereas lung ultrasound (LUS) is increasingly applied at the bedside. Their comparative accuracy vs computed tomography (CT), the reference standard, is uncertain.

Methods We performed a systematic review and Bayesian hierarchical network meta-analysis of prospective diagnostic accuracy studies. PubMed, CINAHL, Scopus, and Web of Science were searched to May 15, 2025. Eligible studies enrolled ICU adults with LUS and/or CXR vs CT. Exclusions were retrospective designs, case series, non-ICU settings, and studies without CT. Risk of bias was assessed with QUADAS-C. Primary outcomes were sensitivity, specificity, log diagnostic odds ratio (logDOR), and hierarchical summary receiver operating characteristic [HSROC] area under the curve (AUC). Bayesian models accounted for heterogeneity. Sensitivity, calibration, and meta-regressions (posterior lung zones and sample size) were conducted. This study is registered with PROSPERO, CRD420251081853.

Findings Of 3080 records, nine studies with 746 patients were included. Risk of bias was generally low for patient selection and flow/timing but high for index test blinding. Grading of Recommendations, Assessment, Development and Evaluation [GRADE] certainty was moderate for LUS and low for CXR. LUS showed higher sensitivity (0.93, 95% credible interval [CrI] 0.91–0.95) than CXR (0.65, 95% CrI 0.62–0.68), comparable specificity (0.83, 95% CrI 0.81–0.85 vs 0.81, 95% CrI 0.79–0.83), higher accuracy (AUC 0.88, 95% CrI 0.77–0.95 vs 0.76, 95% CrI 0.60–0.88), and greater logDOR (4.17, 95% CrI 2.53–6.64 vs 1.43, 95% CrI 0.22–2.86). Between-study heterogeneity was moderate to substantial. For LUS, τ was 0.99 (95% CrI 0.23–3.04) for sensitivity, 0.68 (95% CrI 0.18–2.06) for specificity, 1.66 (95% CrI 1.08–2.84) for logDOR, and 0.086 (95% CrI 0.046–0.156) for AUC. For CXR, τ was 0.89 (95% CrI 0.25–2.43) for sensitivity, 0.26 (95% CrI 0.05–1.05) for specificity, 2.32 (95% CrI 1.39–4.77) for logDOR, and 0.175 (95% CrI 0.117–0.231) for AUC.

Interpretation LUS shows higher sensitivity and accuracy than CXR in ICU pneumonia, particularly when dorsal zones are assessed. Moderate specificity, operator dependence, and bias require caution in interpreting the results. Randomised trials should determine whether broader LUS use improves antibiotic stewardship and ICU outcomes.

Funding There was no funding source for this study. Open access publication fees are supported by institutional funds from the University of Udine.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Lung ultrasound; Chest X-ray; Bayesian meta-analysis; Ventilator-associated pneumonia; Diagnostic accuracy

Introduction

Lung ultrasound (LUS) has become an increasingly popular diagnostic tool for pneumonia, providing multiple advantages over traditional imaging modalities, particularly in critically ill patients. Several meta-analyses have demonstrated that LUS has high

*Corresponding author. Department of Emergency, Udine University Hospital, Azienda Sanitaria Universitaria Friuli Centrale (ASU FC), P.le S. Maria della Misericordia, 15, 33100, Udine, Italy.
E-mail address: daniele.orso@asufc.sanita.fvg.it (D. Orso).

SHOUT
SHARE
IT

Research in context

Evidence before this study

We systematically searched PubMed, Embase, Scopus, and the Cochrane Central Register of Controlled Trials (CENTRAL) for studies comparing the diagnostic accuracy of lung ultrasound (LUS) and chest X-ray (CXR) against computed tomography (CT) as the reference standard for pneumonia in critically ill adult patients admitted to the intensive care unit (ICU). The search included all articles published from database inception to May 15, 2025, without language restrictions. Search terms included combinations of "lung ultrasound", "chest X-ray", "computed tomography", "pneumonia", "diagnostic accuracy", "ICU", and "critical care". Additional references were identified by manual screening of bibliographies of relevant reviews. We included studies that assessed LUS or CXR against CT in adult ICU patients with suspected pneumonia. Pediatric studies or those without stratified adult ICU data were excluded. We identified nine eligible studies enrolling a total of 746 critically ill patients. Most were observational, with high or unclear risk of bias due to lack of blinding and inconsistent evaluation of posterior lung zones. Previous meta-analyses have not conducted a formal network meta-analysis under a Bayesian framework in ICU populations with CT as the common reference standard.

Added value of this study

To our knowledge, this is the first Bayesian hierarchical network meta-analysis comparing LUS and CXR for pneumonia diagnosis specifically in ICU patients, using CT as the reference standard. This approach enabled us to integrate direct and indirect evidence under a unified Bayesian framework and to explore methodological sources of heterogeneity. We observed that LUS consistently outperformed CXR in terms of diagnostic accuracy. Meta-regression suggested that omission of dorsal lung zones may reduce diagnostic performance. Calibration analyses and Bayesian assessments of small-study effects supported the robustness of the findings.

Implications of all the available evidence

In the context of critically ill ICU patients, LUS appears to offer a more sensitive and clinically useful alternative to CXR for pneumonia diagnosis, while maintaining comparable specificity. These findings, together with the practical advantages of LUS at the bedside, support its consideration as a valuable diagnostic tool in this setting. However, further studies are needed to assess its impact on clinical management and outcomes, and to better clarify its role in pneumonia subtypes (CAP, HAP, VAP) across diverse ICU populations.

sensitivity/specificity for diagnosing pneumonia in adults[1–3] and performs particularly well in settings such as emergency departments (ED).[3,4]

However, the diagnostic landscape significantly changes in the intensive care unit (ICU). The conditions of patients who are not ambulatory, frequently sedated, intubated, or on non-invasive ventilation have a significant impact on both image acquisition and interpretation. Portable chest X-rays (CXRs), despite being widely used, suffer from limitations in image quality and positioning, resulting in decreased sensitivity compared to computed tomography (CT) scans.[5,6] Some studies have highlighted the reduced diagnostic accuracy of CXR in the ICU, with particularly poor sensitivity in early or subtle cases.[7,8]

LUS is an alternative option that can be carried out at the bedside, avoids radiation, and allows for real-time dynamic evaluation, which is crucial for monitoring changes over time. It has been suggested that LUS can still perform well in the ICU, but the performance may differ from ED settings.[9,10]

Despite this growing body of evidence, few studies have directly compared the diagnostic accuracy of LUS and CXR for pneumonia in critically ill adults admitted to the ICU.[7] The aggregation of data from heterogeneous clinical environments is a limitation of many existing meta-analyses, which limits generalizability to ICU patients. A focused diagnostic network meta-analysis comparing LUS and CXR in ICU patients is necessary due to the unique clinical and technical challenges in intensive care. This approach enables the integration of direct and indirect comparisons of treatment efficacy while accounting for study-level heterogeneity, in order to support evidence-based decision-making in this high-risk patient population.

The present review specifically addressed pneumonia, as an umbrella term encompassing hospital-acquired pneumonia (HAP), ventilator-associated pneumonia (VAP), and severe community-acquired pneumonia (severe-CAP), as the target condition for diagnostic accuracy evaluation in critically ill adults. The aim of this study was to evaluate, using a Bayesian network meta-analysis, the comparative diagnostic accuracy of LUS and CXR vs CT in the diagnosis of pneumonia in ICU patients.

Methods

Search strategy and selection criteria

A systematic review with a Bayesian network diagnostic meta-analysis was performed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines.[11] The systematic review's protocol was prospectively registered in the International Prospective Register of Systematic

Reviews (PROSPERO 2025: CRD420251081853. Available from https://www.crd.york.ac.uk/PROSPERO/view/CRD420251081853).

We included all prospective diagnostic accuracy studies comparing LUS and/or CXR against CT as the reference standard for the diagnosis of pneumonia in adult patients admitted in the ICU setting. For the review's purposes, "pneumonia" was intended as an umbrella term including hospital-acquired pneumonia (HAP), ventilator-associated pneumonia (VAP), and severe community-acquired pneumonia (CAP) treated in the ICU.

We included three types of comparisons:

- LUS vs CT;
- CXR vs CT;
- LUS vs CXR vs CT (triangular comparison with CT as the common reference standard).

Studies were considered eligible if they reported sufficient data to construct 2 × 2 contingency tables (true positives, false positives, true negatives, false negatives). Studies were also excluded if they: (a) did not use CT as the reference standard; (b) included only paediatric populations or mixed cohorts without stratified adult data; (c) were retrospective studies, case reports/series, editorials, reviews, or conference abstracts without full text; (d) were in other languages than English. Sex and/or gender of participants were reported as available in the included studies. When data were not stratified by sex or gender, we extracted overall study-level results. No reclassification of sex or gender was attempted.

We systematically searched the biomedical databases PubMed, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, and Web of Science from inception to May 15, 2025. Search strings were initially developed for the PubMed database by one researcher (FF), and consecutively peer-reviewed by a senior researcher (DO) following the Peer Review of Electronic Search Strategies (PRESS) guidelines checklist.[12] The terms "lung ultrasound", "thoracic ultrasound", "chest x-ray", "pneumonia", and "intensive care unit" were combined using the AND/OR Boolean operators. We included the term "pneumonitis" to capture a broader spectrum of radiological terminology potentially applied to infectious lung infiltrates in ICU settings. During screening, we excluded studies that clearly focused on non-infectious inflammatory conditions. Search strings were translated for the CINAHL, Scopus and Web of Science databases using the Polyglot Search Translator[13,14] and consecutively adapted. No restrictions were applied to the search queries. The complete search strings are available in Table S1 in the Supplementary Material.

Search results were imported into the Covidence platform by Veritas Health Innovation Ltd,[15] licensed to the Residency Program in Anaesthesia and Intensive Care Medicine of the University of Udine. Duplicates were automatically removed by the platform. All the remaining records were independently screened by two reviewers (DO, FF) in two distinct phases: (a) title/abstract screening and (b) full-text assessment. Any disagreement was resolved by discussion until consensus was reached, involving a third researcher (TB). Data were extracted in duplicate using a pre-tested standardised form which included data regarding the study design, population, setting, index tests, diagnostic target condition, and contingency data.

Data analysis

The quality of each study was assessed using the Quality Assessment of Diagnostic Accuracy Studies–Comparative [QUADAS-C] tool.[16] Two reviewers (DO, TB) evaluated risk of bias and applicability concerns across four domains. Also in this phase, any disagreements were resolved through discussion, with a third researcher available for consultation (FF).

To synthesize comparative accuracy across tests, we implemented a Bayesian network meta-analysis (NMA) using a hierarchical model for diagnostic test accuracy. Posterior estimates for logit-transformed sensitivity and specificity were generated via Markov chain Monte Carlo (MCMC) sampling in JAGS (Just Another Gibbs Sampler) through the rjags package in R,[17] using three chains of 30,000 iterations each (10,000 burn-in, thinning = 1); convergence was confirmed by trace plots and Gelman–Rubin statistics ($\hat{R} < 1.05$) and effective sample size. We employed Normal (0, 0.01) priors on the logit scale for sensitivity and specificity parameters. Sensitivity analyses using broader priors (Normal [0,1]) and prior predictive checks confirmed the robustness of estimates. Between-study heterogeneity was estimated using standard deviation parameters on the logit scale (τ).

We applied a bivariate random-effects model and a hierarchical summary receiver operating characteristic (HSROC) model to estimate pooled sensitivity, specificity, likelihood ratios (LHR), and logarithmic diagnostic odds ratio (logDOR), with corresponding 95% credibility intervals. Although data were categorized by comparison type (LUS vs CT, CXR vs CT), all analyses were conducted using a hierarchical Bayesian network meta-analytic framework integrating both direct and indirect comparisons. Stratified analyses by comparison type were not feasible due to the absence of direct head-to-head studies. We estimated positive and negative predictive values (PPV and NPV) assuming a fixed pre-test probability of 40%, as used in the Fagan nomogram analysis. This approach allows indirect estimation of post-test probabilities in a clinically relevant setting.

We calculated Surface Under the Cumulative Ranking Curve [SUCRA] values for each diagnostic metric (sensitivity, specificity, area under the curve

[AUC], and logDOR). Rankings were derived from the cumulative distribution of each test's performance across posterior samples.

To further enhance the capability of the meta-analysis to translate the diagnostic accuracy into clinical decision-making, we applied Fagan nomograms considering VAP, severe CAP and HAP, given their high clinical relevance in the ICU, and using a pre-test probability consistent with published estimates of the prevalence of VAP, severe CAP and HAP in ICU patients.[18–21]

Leave-one-out sensitivity analyses and influence diagnostics were performed to assess the robustness of pooled estimates. We assessed potential publication bias through visual inspection of funnel plots and performed Egger's regression test to quantify small-study effects. A p-value <0.10 on Egger's test was considered indicative of asymmetry and potential publication bias. These analyses were conducted separately for sensitivity, specificity, and logDOR estimates.

Two separate meta-regression models were run to evaluate the effect of (i) sample size, and (ii) the omission of dorsal lung zones by the individual studies, respectively. For each model, the slope coefficient represents the estimated change in logDOR associated with the covariate of interest.

Statistical analyses were performed in R version 4.5.0 (R Foundation for Statistical Computing, Vienna, Austria). Bayesian hierarchical network meta-analyses of diagnostic test accuracy were implemented in JAGS (via the rjags package in R) using a hierarchical summary receiver operating characteristic (HSROC) framework. Additional Bayesian regression and calibration models were conducted using Stan (via the brms and rstanarm/stan_glm packages), while funnel plots and frequentist Egger's tests were generated with base R functions.

Role of the funding source

There was no funding source for this study. DO and FF have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Both authors are responsible for the decision to submit the manuscript for publication.

Results

A total of 3080 records were screened after removal of duplicates, of which nine studies comprising 746 critically ill adult patients met the inclusion criteria for the Bayesian hierarchical diagnostic network meta-analysis[22–30] (Fig. 1). All studies compared LUS and/or CXR against CT scan, which served as the reference standard[22–30] (Table 1). The studies were conducted across multiple countries, including France, Egypt, Kuwait, and China, and primarily involved patients with suspected pneumonia (i.e., CAP, VAP or HAP) or suspected VAP. Most studies used consecutive sampling and were single-centred in design. None of the studies reported blinded interpretation of index tests. Operator expertise varied, with LUS being performed by intensivists alone or in combination with radiologists (Table S2 in the Supplementary Material). The scanning of dorsal lung zones was inconsistently reported (3 studies), and in some studies explicitly omitted (3 studies). Definitions of pneumonia diagnosis via LUS and CXR were heterogeneous across studies, ranging from alveolar consolidation patterns, dynamic air bronchograms and B-lines for LUS, to lobar opacities or infiltrates on CXR, though definitions were not fully standardized across studies (Table 1). Among the included studies, 8 arms evaluated the diagnostic performance of LUS against CT, and 6 arms assessed CXR against CT; however, no study directly compared LUS and CXR within the same population (Figure S1 in the Supplementary Material). Among included studies, most did not stratify diagnostic performance by sex or gender; therefore, our pooled analyses represent overall populations without sex- or gender-specific estimates.

Most studies (6/9 studies; 67%) demonstrated low risk of bias in the patient selection domain, suggesting appropriate inclusion criteria and enrolment procedures (Table 2). However, a consistent concern across multiple studies (8/9 studies; 89%) was the absence of blinding between index tests and reference standards, resulting in frequent high-risk assessments for the index test domain. Furthermore, several studies (8/9 studies; 89%) did not explicitly state whether the reference standard was interpreted independently, leading to "unclear" judgments in that domain. Flow and timing were generally well documented, with many studies performing diagnostic tests within a 24-h window (7/9 studies; 78%), supporting low risk assessments. Overall, while the included studies provide valuable diagnostic insights, methodological improvements in blinding procedures would significantly enhance internal validity (Figure S2 in the Supplementary Material).

LUS demonstrated superior diagnostic accuracy with a posterior sensitivity of 0.93 (95% credible interval [CrI] 0.91–0.95), specificity of 0.83 (95% CrI 0.81–0.85), and an area under the HSROC curve (AUC) of 0.88 (95% CrI: 0.77–0.95) (Fig. 2 and Figure S3 in the Supplementary Material). In contrast, CXR showed lower sensitivity at 0.65 (95% CrI 0.62–0.68), comparable specificity at 0.81 (95% CrI 0.79–0.83), and an AUC of 0.76 (95% CrI 0.60–0.88). LUS showed an LHR^+ of 5.45 (95% CrI 4.72–6.28), LHR^- of 0.08 (95% CrI: 0.06–0.11) and a logDOR of 4.17 (95% CrI 2.53–6.64). In comparison, CXR showed an LHR^+ of 3.42 (95% CrI 2.81–4.11), LHR^- of 0.43 (95% CrI 0.35–0.53), and a logDOR of 1.43 (95% CrI 0.22–2.86). Assuming a pre-test probability of 40%, the posterior

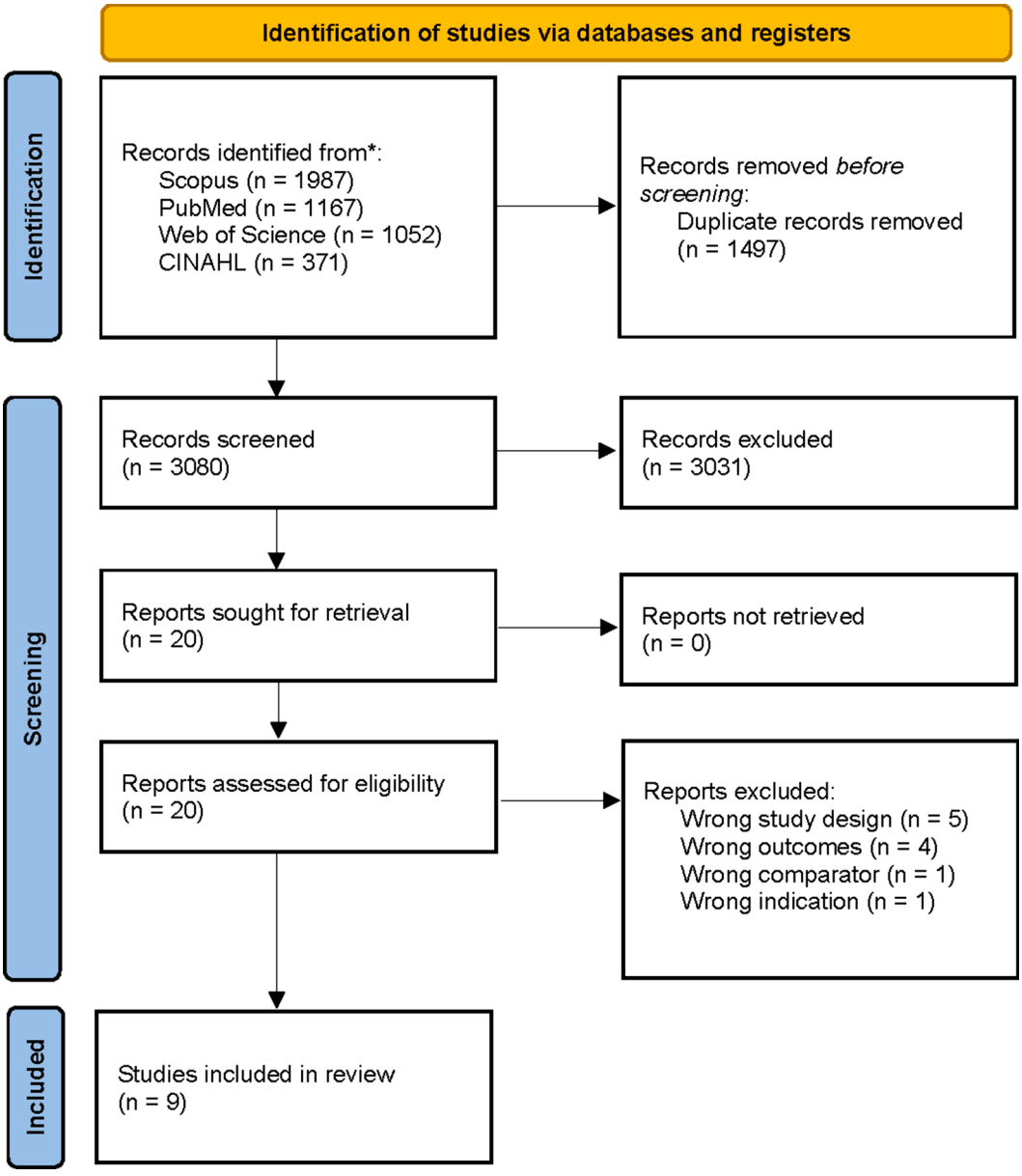

***Fig. 1:* PRISMA flow diagram of study selection**. Flow chart summarising the identification, screening, eligibility assessment, and inclusion of studies in the diagnostic network meta-analysis comparing lung ultrasound (LUS) and chest X-ray (CXR) for pneumonia detection in intensive care unit (ICU) patients. Of the several thousand records initially retrieved, nine studies met the predefined eligibility criteria and were included in the final analysis.

predictive values were: for LUS, PPV 0.78 (95% CrI 0.75–0.82) and NPV 0.96 (95% CrI 0.94–0.97); for CXR, PPV 0.71 (95% CrI 0.68–0.75) and NPV 0.87 (95% CrI 0.83–0.89). Posterior HSROC plots revealed minimal overlap between LUS and CXR AUC distributions (Fig. 3), visually confirming the superior accuracy of LUS. All monitored parameters showed excellent convergence, with R-hat values of 1.00. The effective sample size (ESS) was adequate for all key parameters, with values ranging from 457 to 7545 (Table S3 in the Supplementary Material).

Posterior estimates of between-study heterogeneity (τ, on the logit scale) were consistent with moderate to substantial variability across studies. For LUS, τ was 0.99 (95% CrI 0.23–3.04) for sensitivity, 0.68 (95% CrI 0.18–2.06) for specificity, 1.66 (95% CrI 1.08–2.84) for logDOR, and 0.086 (95% CrI 0.046–0.156) for AUC. For CXR, τ was 0.89 (95% CrI 0.25–2.43) for sensitivity,

First author	Country	Inclusion criteria	Index tests evaluated	Operator	Sample size	Posterior field explored	Diagnostic criteria
Beydon_1992	France	Post-surgery acute hypoxic respiratory failure	CXR	na	19	na	Consolidation
Elmahalawy_2017	Egypt	Suspected pneumonia	CXR, LUS	intensivist	130	no	Consolidation
Bitar_2018	Kuwait	Suspected pneumonia	CXR, LUS	intensivist	38	yes	Consolidation + Air bronchogram
Haggag_2019	Egypt	Suspected CAP	CXR, LUS	intensivist	100	na	Consolidation + Air bronchogram
Zhou_2019	China	Suspected VAP	LUS	intensivist	124	no	Consolidation + Air bronchogram
Gaber_2020	Egypt	Suspected VAP	CXR, LUS	intensivist	80	na	na
Haaksma_2021	Netherlands	Suspected pneumonia	LUS	researcher	120	yes	Consolidation + Air bronchogram
Dhawan_2022	India	Suspected pneumonia	CXR, LUS	intensivist	85	yes	Consolidation + Air bronchogram
Babu_2025	India	Suspected VAP	LUS	resident	50	no	LUS score

Summary of the nine studies comparing lung ultrasound (LUS) and chest X-ray (CXR) for pneumonia diagnosis in critically ill patients. The table reports country of origin, inclusion criteria, index tests evaluated, operator expertise, and sample size. Posterior field exploration indicates whether dorsal lung zones were systematically assessed. Diagnostic criteria (e.g., consolidation, air bronchogram), where available, are also shown. When studies refer to "pneumonia," this typically encompasses a mix of community-acquired pneumonia (CAP), hospital-acquired pneumonia (HAP), or ventilator-associated pneumonia (VAP), without systematic distinction. Abbreviations: ICU, intensive care unit; LUS, lung ultrasound; CXR, chest X-ray; AUC, area under the ROC curve; na, not available.

Table 1: **Characteristics of included studies.**

0.26 (95% CrI 0.05–1.05) for specificity, 2.32 (95% CrI 1.39–4.77) for logDOR, and 0.175 (95% CrI 0.117–0.231) for AUC.

A prior sensitivity analysis comparing weakly informative Normal (0, 0.01) and broader Normal (0, 1) priors revealed minimal qualitative differences in the ranking of LUS vs CXR, although some posterior estimates shifted slightly. Notably, LUS sensitivity and specificity increased under broader priors, while CXR estimates remained less stable (Table S5 in the Supplementary Material).

SUCRA-based rankings consistently favoured LUS across all key metrics (Figure S4 in the Supplementary Material). The SUCRA analysis for logDOR confirmed that LUS ranked first in virtually all posterior samples. The estimated SUCRA was 99.9% (95% CrI 99.7%–100.0%) for LUS and 0.10% (95% CrI 0.00%–0.30%) for CXR, indicating that LUS consistently outperformed CXR in overall diagnostic efficiency (Tables S4 and S5 in the Supplementary Material).

Calibration plots showed close agreement between model-predicted and observed values (Figures S5–S10 in the Supplementary Material). For LUS, the calibration slope for sensitivity was 1.47 (95% CrI 1.33–1.61) with an intercept of –0.44 (95% CrI –0.57 to –0.31), indicating a modest, systematic under-prediction of study-level sensitivity. The specificity slope was 1.19 (95% CrI 0.76–1.62) with an intercept of –0.15 (95% CrI –0.52 to 0.22); although the point estimate also suggested mild under-prediction, the wide CrI denotes substantial uncertainty. For CXR, calibration was satisfactory: the sensitivity slope was 1.10 (95% CrI 1.04–1.15) and intercept –0.06 (95% CrI –0.10 to –0.02), while the specificity slope was 1.24 (95% CrI 0.81–1.63) with an intercept of –0.21 (95% CrI –0.47 to 0.09). Taken together, these analyses indicate that the HSROC model slightly underestimates the accuracy of LUS, particularly its sensitivity, whereas predictions for CXR are well aligned with the observed data.

Because no study reported a direct comparison between LUS and CXR, inconsistency could be assessed only at the global level with the design-by-treatment interaction. This test showed significant incoherence ($Q = 8.51$; $df = 3$; $p = 0.037$). Leave-one-out analysis indicated that removing the study by Elmahalawy et al. reduced the statistic to $Q = 4.69$ ($df = 2$; $p = 0.096$) and removing the study by Bitar et al. reduced it further to $Q = 2.28$ ($df = 2$; $p = 0.32$); exclusion of any other single study left the network inconsistent ($p < 0.05$). Thus, the trials by Elmahalawy and by Bitar each account for the observed incoherence. Updated pooled estimates of sensitivity, specificity, AUC, and logDOR after omitting

	Test comparison	Patient selection	Index test bias	Reference standard bias	Flow and timing
Babu_2025	LUS vs CT	Unclear	High risk	Unclear	Unclear
Beydon_1992	CXR vs CT	Unclear	High risk	Unclear	Unclear
Bitar_2018	CXR vs CT	Low risk	High risk	Unclear	Low risk
Bitar_2018	LUS vs CT	Low risk	High risk	Unclear	Low risk
Dhawan_2022	CXR vs CT	Unclear	High risk	Unclear	Low risk
Dhawan_2022	LUS vs CT	Unclear	High risk	Unclear	Low risk
Elmahalawy_2017	CXR vs CT	Low risk	High risk	Unclear	Low risk
Elmahalawy_2017	LUS vs CT	Low risk	High risk	Unclear	Low risk
Gaber_2020	CXR vs CT	Low risk	High risk	Unclear	Low risk
Gaber_2020	LUS vs CT	Low risk	High risk	Unclear	Low risk
Haaksma_2021	LUS vs CT	Low risk	Low risk	Unclear	Low risk
Haggag_2019	CXR vs CT	Low risk	High risk	Unclear	Low risk
Haggag_2019	LUS vs CT	Low risk	High risk	Unclear	Low risk
Zhou_2019	LUS vs CT	Low risk	High risk	Low risk	Low risk

The risk of bias assessment for each diagnostic accuracy study included in the network meta-analysis, using the QUADAS-C tool. The evaluation covered four domains: patient selection, index test bias, reference standard bias, and flow and timing. Each domain was graded as 'Low risk', 'High risk', or 'Unclear', according to the criteria set by the QUADAS-C framework. A study was considered at high risk of bias if at least one domain was judged to be at high risk. LUS: lung ultrasound; CXR: Chest X-ray; CT: Chest Tomography.

Table 2: **Risk of bias assessment using QUADAS-C.**

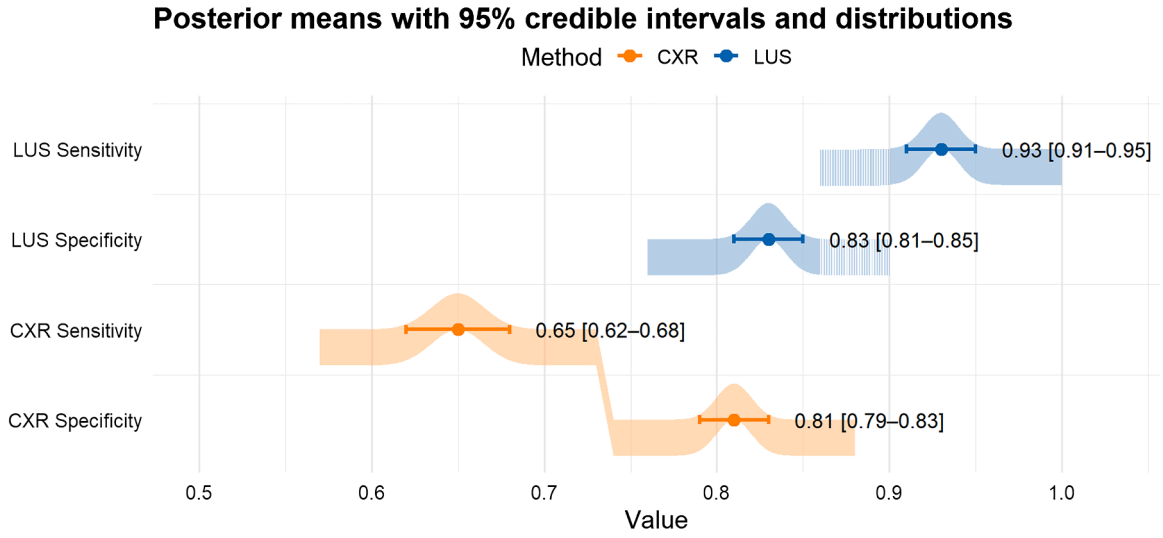

Fig. 2: **Posterior distributions of sensitivity and specificity for LUS and CXR**. Posterior distributions and 95% credible intervals for sensitivity and specificity of lung ultrasound (LUS) and chest X-ray (CXR) in the diagnosis of pneumonia in ICU patients. Coloured densities (light blue for LUS, orange for CXR) represent the Bayesian posterior distributions. Dots indicate posterior means, with horizontal bars denoting 95% credible intervals.

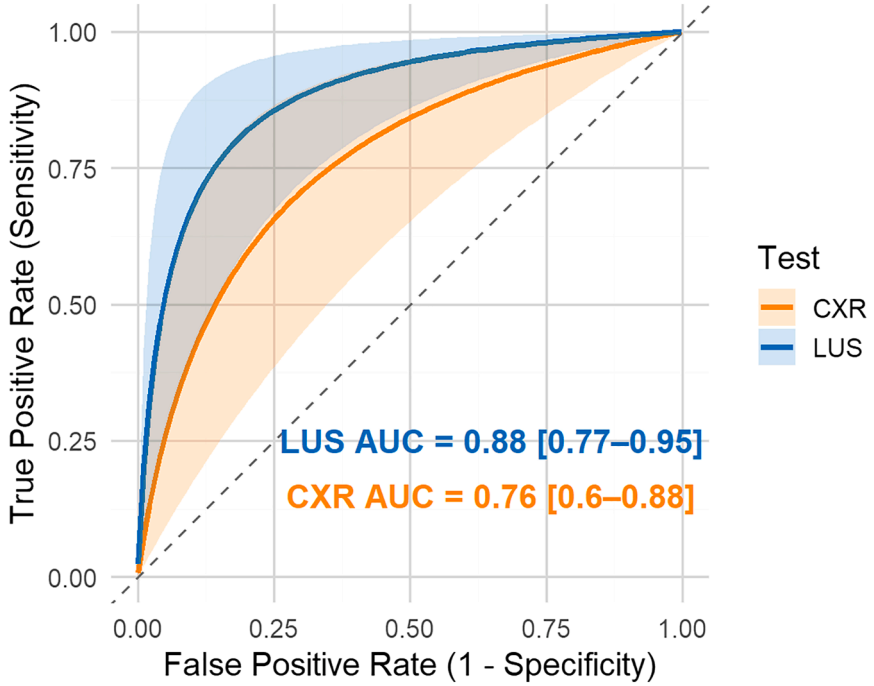

Fig. 3: **Summary HSROC curves with 95% credible intervals for LUS and CXR**. Hierarchical summary receiver operating characteristic (HSROC) curves comparing lung ultrasound (LUS) and chest X-ray (CXR) for the diagnosis of pneumonia in ICU patients. Solid lines represent the mean posterior curves for each test, while shaded areas denote the 95% Bayesian credible regions. No individual study estimates are shown. LUS demonstrated higher diagnostic accuracy with an area under the curve (AUC) of 0.88 (95% CrI: 0.77–0.95), compared to 0.76 (95% CrI: 0.60–0.88) for CXR.

these two studies are provided in Supplementary Table S6.

Bayesian Egger's regression revealed strong asymmetry for both index tests (Fig. 4). The slope was −25.2 (95% CrI −25.35 to −25.05) for LUS and −25.75 (95% CrI −25.98 to −25.52) for CXR. The Bayesian p-value for the null hypothesis of slope ≥0 was <0.00001 in both cases, suggesting substantial small-study effects. The posterior distribution of the precision coefficient (Figure S11 in the Supplementary Material) further supported these findings, showing a markedly asymmetric and narrow distribution centred far below zero for both index tests, reinforcing the evidence of small-study effects.

The probability of post-test was shown through Fagan's nomograms for VAP, severe CAP, and HAP (Fig. 5). In all three clinical scenarios, LUS consistently produced a greater shift in diagnostic probability compared to CXR. For VAP, with an assumed pre-test probability of 20%, LUS increased the post-test probability to 58% after a positive result and reduced it to 1.8% after a negative result. In contrast, CXR yielded a post-test probability of 46% (positive) and 9.8% (negative). In severe CAP requiring ICU admission (pre-test 50%), LUS resulted in a post-test probability of 85% with a positive test and 7.7% with a negative test, while CXR yielded 77% and 30%, respectively. For HAP in ICU (pre-test 30%), LUS generated post-test probabilities of 70% (positive) and 3% (negative), compared to 58% and 16% with CXR.

Two univariable meta-regressions were conducted, both using logDOR as the outcome. The first assessed whether study sample size influenced diagnostic performance and found no consistent association (posterior slope = 0.75; 95% CrI −0.92 to 2.40). The second evaluated whether omission of dorsal lung zones affected accuracy; studies that did not assess dorsal lung zones showed a lower logDOR (β = −3.55; 95% CrI −7.14 to 0.07), although this association did not reach statistical credibility. These analyses are illustrated in Figure S12.

Based on GRADE criteria,[31] the certainty of evidence was rated as moderate for LUS and low for CXR (Table 3). The diagnostic accuracy of LUS was supported by high sensitivity and good specificity, but the overall rating was downgraded due to risk of bias (primarily related to dorsal lung zones protocols and lack of blinding) and some concerns about indirectness (Table S7 in the Supplementary Material). For CXR, the evidence was downgraded for imprecision—given the relatively wide credible intervals—and inconsistency in sensitivity estimates across studies.

Discussion

Our Bayesian network meta-analysis shows that, in critically ill patients, LUS outperforms CXR in terms of sensitivity for pneumonia diagnosis, while maintaining comparable specificity. The inclusion of dorsal lung zones could affect diagnostic performance. LUS was

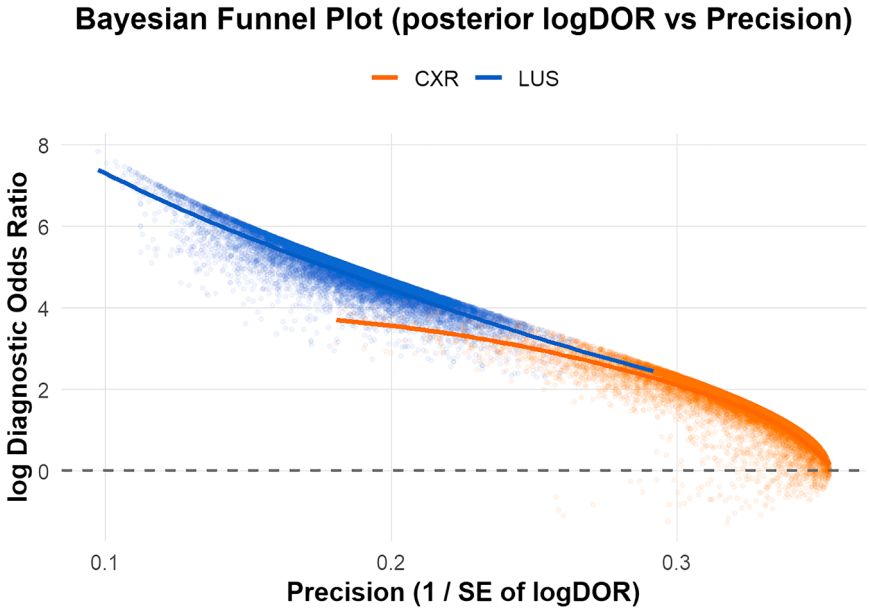

Fig. 4: **Bayesian funnel plots assessing small-study effects for LUS and CXR**. Bayesian funnel plots showing the relationship between study precision and log diagnostic odds ratio (logDOR) for LUS and CXR. Visual asymmetry was observed for LUS, consistent with potential small-study effects or publication bias. Dots represent individual studies; lines correspond to posterior regression estimates.

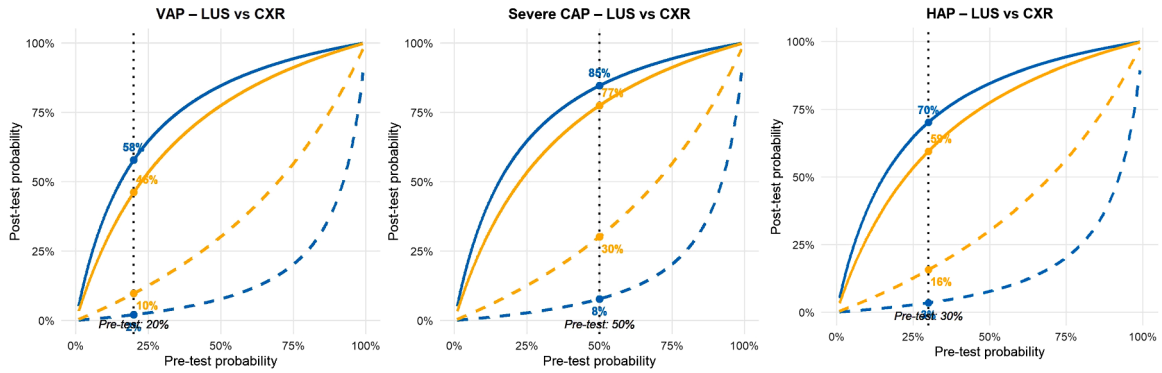

Fig. 5: **Fagan nomograms based on posterior estimates for LUS and CXR**. Each panel shows a Fagan nomogram comparing lung ultrasound (LUS, blue) and chest X-ray (CXR, orange) for pneumonia diagnosis in ICU patients with VAP (top-left), severe CAP (top-right), and HAP (bottom-left). Solid lines represent post-test probability after a positive result; dashed lines after a negative result. Vertical dotted lines indicate pre-test probabilities based on available epidemiologic estimates (20% for VAP, 50% for severe CAP, 30% for HAP). Posterior probabilities are derived from the Bayesian HSROC model.

able to detect approximately 12–25% more cases of pneumonia than CXR when using chest CT as the reference standard.[32] Correspondingly, the posterior estimates from our analysis indicate that LUS offers substantially higher sensitivity (on the order of ~90% vs ~60% for CXR) while maintaining similar or only slightly higher specificity. This translates into a better overall diagnostic performance for LUS confirmed by the literature: in one large series the area under the receiver operating characteristic curve was 0.97 for LUS compared to 0.87 for CXR.[10] Consistently, LUS ranked as the top diagnostic modality in our network meta-analysis (with a nearly maximal SUCRA value, indicating the highest probability of being the best test), far outpacing the performance of CXR.

As shown in Figure S13, our pooled AUC estimate for LUS aligns closely with prior meta-analyses, reinforcing its high overall diagnostic accuracy. In contrast, the AUC for CXR in our analysis is substantially lower, largely due to decreased specificity. These differences may be attributed to stricter inclusion criteria, the use of CT as a consistent reference standard, and a Bayesian hierarchical modelling approach that accounts for study-level heterogeneity. Compared to previous meta-analyses conducted in ED or general acute settings—such as those by Chávez et al.,[2] Long et al.,[3] and Orso et al.[4]—our analysis reports slightly lower specificity for LUS. For example, Orso et al. reported a pooled specificity of 90%, while our results are lower.[4] This discrepancy likely reflects the greater diagnostic challenge in ICU patients, where posterior lung zones are often not fully assessed due to patient instability, supine positioning, or mechanical ventilation. Our meta-regression further supports this explanation, showing a trend toward reduced diagnostic accuracy in studies that omitted dorsal lung zones evaluation. These methodological differences highlight the importance of setting-specific evaluations and may explain why LUS performs differently in ICU vs ED populations.

In the ICU, diagnostic imaging is constrained by patient instability and limited access to advanced scanners. Bedside CXR is usually restricted to a single supine projection, resulting in sub-optimal sensitivity, whereas transporting patients receiving mechanical ventilation to CT carries logistic cost, radiation exposure, and a well-documented risk of cardiorespiratory deterioration.[33,34] LUS can be performed at the bedside in minutes, requires no ionising radiation, and provides dynamic, repeatable information that is immediately actionable.[2] Our findings therefore suggest that LUS may be considered the preferred first-line imaging test when pneumonia is suspected in critically ill adults, reserving CT for equivocal or complex cases.

Hansell and colleagues performed a meta-analysis in ICU patients that compared LUS and CXR for the generic finding of lung consolidation, irrespective of aetiology.[35] Their pooled estimates (sensitivity 0.91 and specificity 0.77) are numerically close to the values we report, but the two reviews draw on mutually exclusive primary studies because of different eligibility criteria. We included only studies in which pneumonia was clinically adjudicated and CT-confirmed, whereas Hansell et al. accepted any parenchymal consolidation. Owing to this definitional gap, their dataset is not shown in Figure S13, which is limited to pneumonia-specific evidence. Nonetheless, both reviews converge on the same message: bedside LUS is substantially more sensitive than portable CXR in the intensive-care setting, emphasising the need to interpret imaging findings within the clinical context.

Despite the well-defined sonographic features of pneumonia, real-time recognition can be challenging for less experienced operators.[36,37] Mechanically ventilated patients have a high incidence of atelectasis,

Test	Sensitivity (95% CrI)	Specificity (95% CrI)	AUC (95% CrI)	logDOR (95% CrI)	LHR+ (95% CrI)	LHR-(95% CrI)	PPV (95% CrI)	NPV (95% CrI)	Certainty of evidence	Reason for downgrade
LUS	0.93 [0.91–0.95]	0.83 [0.81–0.85]	0.88 [0.77–0.95]	4.17 [2.53–6.64]	5.45 [4.72–6.28]	0.08 [0.06–0.11]	0.78 [0.75–0.82]	0.96 [0.94–0.97]	●●◯ (Moderate)	Risk of bias (posterior), indirectness
CXR	0.65 [0.62–0.68]	0.81 [0.79–0.83]	0.76 [0.60–0.88]	1.43 [0.22–2.86]	3.42 [2.81–4.11]	0.43 [0.38–0.49]	0.71 [0.68–0.75]	0.87 [0.83–0.89]	●◯◯ (Low)	Imprecision, inconsistency

Lung ultrasound (LUS) shows high sensitivity and good specificity in identifying the target condition. Chest X-ray (CXR) demonstrates moderate diagnostic performance, particularly with lower sensitivity. Computed tomography (CT) was used as the reference standard across all included studies.

Table 3: **GRADE summary of findings — diagnostic accuracy NMA.**

which is both frequent and multifactorial.[38] Extrinsic compression caused by prolonged supine positioning or obstructive causes like mucus plugging or impaired bronchial clearance can cause atelectasis. The distinction between pneumonia and atelectasis is based on the presence of dynamic air bronchograms (associated with alveolar filling) vs static or fluid bronchograms (more indicative of passive collapse or obstruction).[28,39,40] This distinction, though diagnostically critical, may be subtle and technically demanding, and it requires sufficient skill in acquiring and interpreting LUS images. The importance of thorough scanning, including dorsal lung zones, and interpretative expertise cannot be underestimated when LUS findings can resemble pneumonia. This is necessary to avoid false positives, which could result in unnecessary antibiotic treatment or delayed recognition of other causes of respiratory failure.

Despite the clear benefits of LUS over CXR in this and other studies, CXR remains deeply entrenched in ICU practice as the first-line imaging for suspected pneumonia.[41] This persistent reliance on CXR, despite its lower diagnostic performance, may stem from both historical practice patterns and a training gap among providers. Many intensivists and acute care physicians have not had formal training in LUS, or lack experienced mentors to develop this skill, as surveys consistently highlight a lack of trained faculty and structured curricula as barriers to LUS adoption.[42] Without sufficient training and confidence in using ultrasound, clinicians may default to the familiarity of CXR even when evidence suggests this one it is inferior to the LUS. Additionally, institutional protocols and ingrained culture change slowly—CXR has been a cornerstone of pneumonia diagnosis for decades, and current guidelines have been slow to catch up with emerging evidence. Notably, some guidelines still do not formally recommend LUS as an alternative to CXR or CT for pneumonia diagnosis.[2] This cautious stance likely reflects the need for more widespread expertise and standardization in LUS.[43]

This meta-analysis has several limitations. First, the evidence base is small, so pooled estimates remain imprecise and vulnerable to small-study effects. Second, all trials were observational and at high or unclear risk of bias, chiefly because blinding of the index tests to the reference standard was rarely reported and dorsal zones were inconsistently scanned. Third, methodological heterogeneity (patient mix, imaging protocols, and non-standardised definitions of pneumonia) may undermine comparability across studies. Specifically, our network meta-analysis assumes transitivity—that is, that patients in studies comparing LUS to CT are sufficiently similar to those comparing CXR to CT. While all studies were conducted in ICU populations using CT as the reference standard, we acknowledge that clinical and methodological differences may limit

exchangeability between study arms. We were unable to perform subtype analyses for VAP, HAP, or severe-CAP pneumonia because stratified data were sparse. In addition, the meta-regression exploring omission of dorsal zones was based on few studies and its association did not reach statistical credibility. Finally, LUS cannot visualise consolidations that spare the pleural surface, although such lesions are uncommon in adult infectious pneumonia and therefore unlikely to have materially altered our conclusions.[32]

However, it is reassuring that the central qualitative finding—the superiority of LUS over CXR—was consistent across diverse studies and analytic methods. In this regard, a key strength of our study lies in the use of a hierarchical Bayesian multilevel framework, which accounts for both within- and between-study variability and mitigates the impact of attribution bias. This approach enhances the robustness of diagnostic accuracy estimates and allows for a more reliable synthesis across heterogeneous studies.

In conclusion, LUS demonstrates substantially higher diagnostic sensitivity and overall accuracy than CXR for pneumonia in critically ill ICU patients, with superiority in logDOR particularly when dorsal lung zones are assessed. Nevertheless, its moderate specificity, operator dependence, and the risk of bias in primary studies mean that these findings should be interpreted cautiously and support, rather than dictate, clinical protocols. Adequately powered randomised trials are still required to determine whether wider adoption of LUS improves antibiotic stewardship and clinical outcomes in the ICU.

Contributors

DO conceived the study, conducted the statistical analyses, developed the Bayesian models, contributed to data interpretation, and drafted the manuscript. FF supervised the analysis, drafted the manuscript, and contributed to data interpretation. TB performed literature review and contributed to revision. DO and FF have access to and verify the underlying study data. All authors meet the ICMJE criteria for authorship: substantial contributions to the conception or design of the work; drafting or revising it critically for important intellectual content; and final approval of the version to be published.

Data sharing statement

Study-level extracted data and statistical code will be available with publication upon reasonable request to the authors, after approval and completion of a data access agreement.

Declaration of interests

All authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2025.103553.

References

1 Lichtenstein DA. Lung ultrasound in the critically ill. *Ann Intensive Care*. 2014;4(1):1.
2 Chavez MA, Shams N, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. *Respir Res*. 2014;15(1):50.
3 Long L, Zhao HT, Zhang ZY, Wang GY, Zhao HL. Lung ultrasound for the diagnosis of pneumonia in adults: a meta-analysis. *Medicine (Baltim)*. 2017;96(3):e5713.
4 Orso D, Guglielmo N, Copetti R. Lung ultrasound in diagnosing pneumonia in the emergency department: a systematic review and meta-analysis. *Eur J Emerg Med*. 2018;25(5):312–321.
5 Ünlüer EE, Karagöz A. Bedside lung ultrasound versus chest X-ray use in the emergency department. *Interv Med Appl Sci*. 2014;6(4):175–177.
6 van den Berk IAH, Lejeune EH, Kanglie MMNP, et al. The yield of chest X-ray or ultra-low-dose chest-CT in emergency department patients suspected of pulmonary infection without respiratory symptoms or signs. *Eur Radiol*. 2023;33(10):7294–7302.
7 Winkler MH, Touw HR, van de Ven PM, Twisk J, Tuinman PR. Diagnostic accuracy of chest radiograph, and when concomitantly studied lung ultrasound, in critically ill patients with respiratory symptoms: a systematic review and meta-analysis. *Crit Care Med*. 2018;46(7):e707–e714.
8 Alzahrani SA, Al-Salamah MA, Al-Madani WH, Elbarbary MA. Systematic review and meta-analysis for the use of ultrasound versus radiology in diagnosing of pneumonia. *Crit Ultrasound J*. 2017;9(1):6.
9 Abid I, Qureshi N, Lategan N, Williams S, Shahid S. Point-of-care lung ultrasound in detecting pneumonia: a systematic review. *Can J Respir Ther*. 2024;60:37–48.
10 Xia Y, Ying Y, Wang S, Li W, Shen H. Effectiveness of lung ultrasonography for diagnosis of pneumonia in adults: a systematic review and meta-analysis. *J Thorac Dis*. 2016;8(10):2822–2831.
11 McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319(4):388–3882396.
12 McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol*. 2016;75:40–46.
13 Clark JM, Sanders S, Carter M, et al. Improving the translation of search strategies using the polyglot search translator: a randomized controlled trial. *J Med Libr Assoc*. 2020;108(2):195–207.
14 Kung J. Polyglot search translator. *J Can Health Libr Assoc*. 2022;43(1) [cited 2025 Jun. 27] Available from: https://journals.library.ualberta.ca/jchla/index.php/jchla/article/view/29600.
15 Veritas Health Innovation. *Covidence systematic review software*. Melbourne, Australia: Veritas Health Innovation; 2025 [cited 2025 Jun. 27]. Available from: https://www.covidence.org.
16 Yang B, Mallett S, Takwoingi Y, et al. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med*. 2021;174(11):1592–1599.
17 R Core Team. *R: a language and environment for statistical computing. Version 4.5.0*. Vienna, Austria: R Foundation for Statistical Computing; 2025 [cited 2025 Jun. 27]. Available from: https://cran.r-project.org.
18 Kalil AC, Metersky ML, Klompas M, et al. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the infectious diseases society of America and the American thoracic society. *Clin Infect Dis*. 2016;63(5):e61–e111. Erratum in: Clin Infect Dis. 2017;64(9):1298. doi: 10.1093/cid/ciw799. Erratum in: Clin Infect Dis. 2017;65(8):1435. doi: 10.1093/cid/cix587. Erratum in: Clin Infect Dis. 2017;65(12):2161 https://doi.org/10.1093/cid/cix759.
19 Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;58(5):377–382.
20 Chalmers JD, Singanayagam A, Akram AR, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. *Thorax*. 2010;65(10):878–883.
21 American Thoracic Society, Infectious Diseases Society of America. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med*. 2005;171(4):388–416.
22 Beydon L, Saada M, Liu N, et al. Can portable chest x-ray examination accurately diagnose lung consolidation after major abdominal surgery? A comparison with computed tomography scan. *Chest*. 1992;102(6):1697–1703.
23 Elmahalawy II, Doha NM, Ebeid OM, Abdel-Hady MA, Saied O. Role of thoracic ultrasound in diagnosis of pulmonary and pleural

diseases in critically ill patients. *Egypt J Chest Dis Tuberc*. 2017;66 (2):261–266.
24 Bitar ZI, Maadarani OS, El-Shably AM, Al-Ajmi MJ. Diagnostic accuracy of chest ultrasound in patients with pneumonia in the intensive care unit: a single-hospital study. *Health Sci Rep*. 2018;2(1):e102.
25 Haggag YI, Mashhour K, Ahmed K, Samir N, Radwan W. Effectiveness of lung ultrasound in comparison with chest X-Ray in diagnosis of lung consolidation. *Open Access Maced J Med Sci*. 2019;7(15):2457–2461.
26 Zhou J, Song J, Gong S, et al. Lung ultrasound combined with procalcitonin for a diagnosis of ventilator-associated pneumonia. *Respir Care*. 2019;64(5):519–527.
27 Gaber S, Tayeh O, Wahab KA, Mohamed N, Essawy T. Early detection of ventilator-associated pneumonia: bedside tools. *Egypt J Crit Care Med*. 2020;7:74–79.
28 Haaksma ME, Smit JM, Heldeweg MLA, et al. Extended lung ultrasound to differentiate between pneumonia and atelectasis in critically ill patients: a diagnostic accuracy study. *Crit Care Med*. 2022;50(5):750–759.
29 Dhawan J, Singh G. Bedside lung ultrasound as an independent tool to diagnose pneumonia in comparison to chest X-ray: an observational prospective study from intensive care units. *Indian J Crit Care Med*. 2022;26(8):920–929.
30 Babu A, Joshi A, Chakraborti A, et al. Utility of chest ultrasound in the diagnosis of ventilator-associated pneumonia in the critical care unit of a tertiary care center: a prospective observational study. *Monaldi Arch Chest Dis*. 2025. https://doi.org/10.4081/monaldi.2025.2974.
31 Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383–394.
32 Boccatonda A, Cocco G, D'Ardes D, et al. Infectious pneumonia and lung ultrasound: a review. *J Clin Med*. 2023;12(4):1402.
33 Graat ME, Hendrikse KA, Spronk PE, Korevaar JC, Stoker J, Schultz MJ. Chest radiography practice in critically ill patients: a postal survey in the Netherlands. *BMC Med Imag*. 2006;6:8.
34 Murata M, Nakagawa N, Kawasaki T, et al. Adverse events during intrahospital transport of critically ill patients: a systematic review and meta-analysis. *Am J Emerg Med*. 2022;52:13–19.
35 Hansell L, Milross M, Delaney A, Tian DH, Ntoumenopoulos G. Lung ultrasound has greater accuracy than conventional respiratory assessment tools for the diagnosis of pleural effusion, lung consolidation and collapse: a systematic review. *J Physiother*. 2021 Jan;67(1):41–48. https://doi.org/10.1016/j.jphys.2020.12.002.
36 Speidel V, Conen A, Gisler V, Fux CA, Haubitz S. Lung assessment with point-of-care ultrasound in respiratory coronavirus disease (COVID-19): a prospective cohort study. *Ultrasound Med Biol*. 2021;47(4):896–901.
37 Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. *Anesthesiology*. 2004;100(1):9–15.
38 Tusman G, Böhm SH, Tempra A, et al. Effects of recruitment maneuver on atelectasis in anesthetized children. *Anesthesiology*. 2003;98(1):14–22.
39 Shah A, Oliva C, Stem C, Cummings EQ. Application of dynamic air bronchograms on lung ultrasound to diagnose pneumonia in undifferentiated respiratory distress. *Respir Med Case Rep*. 2022;39: 101706.
40 Lichtenstein D, Mezière G, Seitz J. The dynamic air bronchogram. A lung ultrasound sign of alveolar consolidation ruling out atelectasis. *Chest*. 2009;135(6):1421–1425.
41 Jain SN, Modi T, Aswani Y, Varma RU. Chest radiography in adult critical care unit: a pictorial review. *Indian J Radiol Imag*. 2019;29 (4):418–425.
42 Brady AK, Spitzer CR, Kelm D, Brosnahan SB, Latifi M, Burkart KM. Pulmonary critical care fellows' use of and self-reported barriers to learning bedside ultrasound during training: results of a national survey. *Chest*. 2021;160(1):231–237.
43 Volpicelli G, Elbarbary M, Blaivas M, et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med*. 2012;38(4):577–591.